

Tworzenie typologii gatunków piśmiennictwa multimedialnego na przykładzie blogów – propozycja metodologiczna¹

MACIEJ MARYL

Porozumiewamy się za pomocą gatunków. Cała komunikacja międzyludzka opiera się na pewnych regularnościach i konwencjach ułatwiających z jednej strony wyrażenie, a z drugiej – zrozumienie danej treści. Jak zauważa kognitywista David Hanauer w artykule syntetyzującym empiryczną wiedzę o procesach lektury poezji: „punktem wyjścia w procesie lektury jest ustalenie gatunku [*genre decision*]. Określone cechy tekstu, cele czytelnicze i kontekst lektury stanowią źródła informacji, które kierują procesem ustalania gatunku [...]. Po ustaleniu gatunku czytanego tekstu zostaje uruchomiona odpowiednia wiedza proceduralna, która kieruje procesem lektury.

¹ Artykuł powstał w ramach projektu „Blog jako nowa forma piśmiennictwa multimedialnego”, sfinansowanego ze środków Narodowego Centrum Nauki przyznanych na podstawie decyzji numer DEC-2011/03/N/HS2/06232. Dziękuję członkom zespołu badawczego – Krzysztofowi Niewiadomskiemu i Maciejowi Kidawie – za wszelkie uwagi, które przyczyniły się do rozwinięcia tej metodologii w trakcie realizacji projektu.

W ramach owej, dostosowanej do tekstu, wiedzy proceduralnej znajdują się oczekiwane efekty lektury, rozplanowanie uwagi jednostki, zasób procesów kognitywnych, które należy uruchomić, by zbudować znaczenie oraz odpowiednie typy wewnętrznej reprezentacji”².

A zatem, jeśli widzimy wiersz, to wiemy, jak go czytać, podobnie jak w przypadku autobiografii, książki kucharskiej czy powieści grozy – rozpoznanie gatunku prowadzi do uruchomienia odpowiednich strategii postępowania z danym tekstem. Konwencjonalność jest wpisana w naturę komunikacji ludzkiej – jak pokazują badania, tam gdzie nie ma konwencji, ludzie szybko wymyślają nowe, wprowadzając systematykę do arbitralnie używanych znaków. W jednym z eksperymentów badani komunikujący się ze sobą wyłącznie poprzez rysunki szybko zbudowali coś na kształt języka polegającego na tworzeniu złożonych kombinacji rysunkowych za pomocą rysunków niższego rzędu³.

Na gatunek pojmowany jako konwencja komunikacji składają się zatem z jednej strony cechy formalne przekazu (tj. określone zasady budowania komunikatu), jak

² D. Hanauer, *What we know about reading poetry: Theoretical positions and empirical research*, w: *The Psychology and Sociology of Literature: In Honor of Elrud Ibsch*, red. D. Schram, G. Steen, Amsterdam – Philadelphia: John Benjamins 2001, s. 108.

³ Zob. C. A. Theisen, J. Oberlander, S. Kirby, *Systematicity and arbitrariness in novel communication systems*, „Interaction Studies” 2010, nr 11 (1), s. 14–32. Tam także przegląd literatury przedmiotu.

i projektowana sytuacja komunikacyjna z drugiej (tj. rola odbiorcy i operacje potrzebne do poprawnego odczytania tekstu)⁴. Gatunki pozwalają nam kategoryzować przekazy i tworzyć modele poznawcze, ułatwiające porozumiewanie się i szybkie przetwarzanie informacji⁵.

Wraz z upowszechnieniem się komunikacji elektronicznej pojawiają się gatunki nowe, stanowiące najczęściej remediację gatunków istniejących w innych mediach. W przypadku piśmiennictwa są to przetworzone formy kultury druku (np. portal internetowy jako remediacja gazety) lub rękopiśmiennej (np. e-mail)⁶. Choć problematyka genologiczna, to jest status gatunkowy czytanych tekstów, nie spędza zapewne snu z oczu czytelników blogów, warto zadać sobie pytanie, jak klasyfikować tę

⁴ Por. B. Witosz, *Genologia lingwistyczna. Zarys problematyki*, Katowice: Wydawnictwo Uniwersytetu Śląskiego 2005, s. 12–13; E. Balcerzan, *W stronę genologii multimedialnej*, „Teksty Drugie” 1999, nr 6, s. 11–14.

⁵ Zob. P. Stockwell, *Poetyka kognitywna. Wprowadzenie*, przeł. A. Skucińska, Kraków: Universitas 2006, s. 43–48.

⁶ Ze względu na metodologiczne ukierunkowanie tego tekstu nie rozwijam w tym miejscu tej problematyki, odsyłając czytelników do innych prac, w których szerzej ją analizowałem: M. Maryl, *Reprint i hipermedialność – dwa kierunki rozwoju literatury ucyfrowionej*, w: *Tekst w sieci, t. 2: Literatura, społeczeństwo, komunikacja*, red. A. Gumkowska, Warszawa: Wydawnictwa Akademickie i Profesjonalne 2009, s. 83–91; tegoż, *Konwergencja i komunikacja. Gatunki wypowiedzi na stronach internetowych pisarzy*, „Zagadnienia Rodzajów Literackich” 2012, nr 55 (110), z. 2, s. 29–52; tegoż, *Blog jako „dziennik elektroniczny”. Analiza genologiczna blogów pisarzy*, „Zagadnienia Rodzajów Literackich” 2013, nr 56 (112), z. 2, s. 87–110.

formę wypowiedzi i jak umieścić ją na mapie znanych nam gatunków, wreszcie – jakie są specyficzne relacje komunikacyjne, ustanawiane przez ten gatunek. Słowem, by odwołać się do przywołanego Hanauera, jakiego typu wiedza proceduralna jest nam potrzebna do przetwarzania blogów.

W początkowym okresie funkcjonowania blogów na gruncie polskim (w drugiej połowie pierwszej dekady XXI w.) problem gatunkowy blogów rozwiązywano najczęściej przez odniesienie ich do literatury dokumentu osobistego, a zwłaszcza dzienników i pamiętników⁷. Jednakże wraz z rozwojem blogosfery oraz pewną profesjonalizacją warsztatu blogerów – związaną między innymi z faktem, iż od jakiegoś czasu na blogach można zarobić – sytuacja się komplikuje. Blogosfera zdaje się dzisiaj korzystać z wielce różnorodnych form gatunkowych, które trudno przyporządkować do jednej tradycji piśmiennictwa.

Zilustrujmy ten problem za pomocą popularnej blogerki Segritty⁸. Sama autorka proponuje kategoryzację swoich wpisów na: „lifestyle”, „kultura”, „męsko-damskie”,

⁷ Zob. A. Gumkowska, *Blogi wobec tradycji diarystycznej. Nowe gatunki w nowych mediach*, w: *Tekst (w) sieci*, t. 1: *Tekst, język, gatunki*, red. D. Ulicka, Warszawa: Wydawnictwa Akademickie i Profesjonalne 2009, s. 231–243; A. Szczepan-Wojnarska, *Blogi jako forma literacka*, „Pamiętnik literacki” 2006, s. 191–201.

⁸ Odsyłam czytelników do zarchiwizowanej wersji witryny z 14 marca 2015, by mogli zapoznać się z opisywanymi przykładami na własną rękę: Segritta, [online:] <https://web.archive.org/web/20150314191513/http://segritta.pl/> [dostęp 10.10.2015].

„moda i uroda”, „dom i wnętrza”, „parenting”. Jest to z jednej strony klasyfikacja tematyczna, z drugiej – sygnalizuje różnorodność gatunkową tekstów publikowanych na blogu. Przyjrzyjmy się wpisom widocznym na stronie głównej: poradnik dla pożyczających pieniądze od znajomych, uwagi wychowawcze, porady kosmetyczne, krótkie opowiadanie, recenzja, opinia na temat filmiku popularnego w sieci. W ramach jednego blogu znajdujemy teksty odsyłające do różnych modeli gatunkowych. Co więcej, sprawę komplikuje fakt, iż niektórzy blogerzy wykorzystują także inne formy niż tekst pisany – jak podcasty czy filmy. Prezentowane tu badanie zrodziło się z potrzeby uporządkowania tej różnorodności.

Typologizacja gatunków piśmiennictwa multimedialnego

O głęboko zakorzenionej w naszych procesach poznawczych potrzebie kategoryzowania pisze socjolog Kenneth D. Bailey w książce poświęconej typologiom i taksonomiom. Powiada on, iż klasyfikacja „stanowi niemalże metodologiczny ekwiwalent elektryczności – korzystamy z niej codziennie, lecz pozostaje dla nas tajemnicza. To jedna z tych rzeczy, z których korzystamy, nie wiedząc zbyttnio, jak działają”⁹. Wymieniając walory klasyfikacji,

⁹ K. D. Bailey, *Typologies and taxonomies: an introduction to classification techniques*, Thousand Oaks: Sage 1994, s. 1.

Bailey podkreśla, iż pozwala nam ona ująć i zrozumieć pole badawcze, dokonać jego wstępnego opisu, zredukować złożoność materiału, zidentyfikować podobieństwa i różnice między elementami, zbadać powiązania między nimi, porównać określone typy oraz nakreślić swoistą topografię problemu¹⁰. Celem niniejszego artykułu jest prezentacja autorskiej metodologii literaturoznawczej analizy tekstów elektronicznych, wypracowanej ramach projektu badawczego „Blog jako nowa forma piśmiennictwa multimedialnego”. Przedstawione tu propozycje metodologiczne można zastosować do konstruowania typologii innych gatunków multimedialnych i formułując je w taki sposób, by czytelnicy mogli z nich skorzystać przy własnych badaniach na innym materiale.

Do podstawowych założeń badania należą: (1) przeświadczenie o ciągłości między światem *offline* i *online*, (2) operowanie na materiale zastanym oraz (3) stosowanie metody badań zintegrowanych, łączącej podejścia ilościowe i jakościowe. Po pierwsze zatem aplikujemy tu postulat metodologiczny najlepiej chyba ostatnio sformułowany przez Richarda Rogersa: badanie obiektów cyfrowych (czy też, jak chce badacz, metod danego medium) prowadzić ma nas do zdobycia wiedzy wykraczającej poza kulturę cyfrową, przełamującej sposoby myślenia narzucane przez te obiekty¹¹. Innymi słowy, badając nowe

¹⁰ Tamże, s. 12–14.

¹¹ R. Rogers, *Digital methods*, Boston: The MIT Press 2013, s. 5.

gatunki, staramy się umieszczać je w szerszej perspektywie zarówno tradycyjnych (nieinternetowych) gatunków, jak i codziennych ludzkich praktyk, w których te teksty biorą udział. Traktujemy zatem internet jako źródło danych, a nie jedynie przedmiot badań¹². Po drugie, skupiamy się na materiale zastanym, czyli tym, co można znaleźć w sieci, w przeciwieństwie do takich materiałów zamawianych przez badaczy jak na przykład wyniki ankiet¹³. Pozwala to na bardziej obiektywne odwzorowanie faktycznie wykorzystywanych pojęć i kategoryzacji. Analizujemy zatem, między innymi, kategorie, którymi posługują się serwisy blogerskie, by zrekonstruować topografię blogów. Po trzecie, wreszcie, realizujemy tu postulat komplementarnej eksploracyjnej analizy danych (CEDA, *Complementary Explorative Data Analysis*), czyli zintegrowanej metodologii badań internetowych: „Chodzi o zastosowanie kombinacji metod badawczych wykorzystywanych zazwyczaj do analizy wyników empirycznych lub interpretacji. Założenie jest takie, że słabe strony poszczególnych metod – ilościowej czy jakościowej – są równoważone silnymi stronami innych”¹⁴.

¹² Por. Tamże, s.19–21.

¹³ Różnicę tę rozwijam w: M. Maryl, *Virtual Communities – Real Readers: New Data in Empirical Studies of Literature*, w: *New Beginnings in the Study of Literature*, red. W. van Peer, J. Auracher, Cambridge: Cambridge Scholars Publications 2008, s. 394. O samej ankiecie pisze więcej w tym tomie Piotr Siuda.

¹⁴ F. Sudweeks, S. J. Simoff, *Complementary Explorative Data Analysis. The Reconciliation of Quantitative and Qualitative Principles*, w: *Doing*

Choć postulat ten na pierwszy rzut oka przypomina podobne koncepcje triangulacji czy metod mieszanych, badacze podkreślają, iż chodzi tu o ściśle powiązanie metod w jednym procesie badawczym – nie tyle przeprowadzenie dwóch badań z wykorzystaniem różnych podejść, ile na przykład ilościowe zestawienie wyników analiz jakościowych (np. kodów uzyskanych na podstawie analizy tekstu)¹⁵. Realizacja tych postulatów pozwala oprzeć badania humanistyczne na solidnych, empirycznych fundamentach, bez rezygnacji z niezbywalnego dla tej dyscypliny komponentu hermeneutycznego.

Zanim przejdę do szczegółowego opisu metodologii, zaprezentuję skrótowy opis badania i najważniejszych wniosków, by naszkicować ogólny kontekst prezentowanych tu metod. Ze względu na charakter tego artykułu skupię się tu głównie na stronie metodologicznej. Czytelnicy zainteresowani szerszym omówieniem wyników mogą sięgnąć po publikację, do których odsyłam w tekście.

Jak już wspomniałem, jednym z podstawowych założeń badania było uznanie bloga za nazwę ogólnej formy piśmiennictwa, w której ramach występują różne gatunki. Blog traktujemy jak termin tego samego rzędu, co na przykład książka – wyznacza pewną technologiczną

Internet research: Critical issues and methods for examining the Net, red. S. Jones, Thousand Oaks: Sage 1998, s. 37.

¹⁵ Por. tamże, s. 39–52.

i komunikacyjną specyfikę przekazu, natomiast poszczególne realizacje mogą się od siebie różnić (jak tomik poezji od encyklopedii)¹⁶. Celem analiz było zatem wyodrębnienie i nazwanie gatunków blogowych. Badanie podzieliliśmy na cztery etapy. Na pierwszym etapie postanowiliśmy przyjrzeć się typologiom zastanym i zobaczyć, jak sami użytkownicy klasyfikują swoją twórczość. Zespół przeanalizował 287 kategorii i podkategorii wykorzystywanych w polskich serwisach blogowych, wyróżniając pięć głównych wyznaczników: tematyczne (np. „Komputery”, „Języki obce”, „Komiksy”, „Czas wolny”, oddające tematykę bloga, 70 procent wszystkich kategorii); komunikacyjne (np. „Profesjonalne”, „Dla facetów”, „Po trzydziestce”, projektujące partnerów komunikacji, 13 procent), formalne (np. „Obrazkowe”, „Videoblogi”, odnoszące się do formy, jak np. obrazki, 6 procent), gatunkowe (wyznaczające gatunek, np. „Opowiadania”, „Plotka”, „Pamiętnik”, 6 procent), oraz oddające perspektywę opisu świata (np. „Mroczne”, „Crazy”, „Odjechane”, 5 procent)¹⁷.

Nie byliśmy zadowoleni z tego wyniku. Szukając gatunków, wpadliśmy *de facto* w pułapkę tematyki bloga, która służy w większości przypadków za główny wyznacznik typologiczny. W niektórych serwisach (np. Bloog.pl) do opisu blogów można wykorzystać kilka kategorii

¹⁶ Szczegółowe omówienie tego zagadnienia w: M. Maryl, K. Niewiadomski, *Blog = Książka? Empiryczne badanie potocznej kategoryzacji blogów*, „Przegląd Humanistyczny” 2013, nr 4, s. 83–85.

¹⁷ Pełny opis wyników: tamże.

(np. „Pamiętniki-Gwiazdy-Muzyka”, „Czas wolny-Hobby-Humor i rozrywka” itp.). Wyszliśmy z założenia, że takie kombinacje – nazwane przez nas syntagmami – mogą same w sobie stanowić gatunki. Wykorzystaliśmy zatem syntagmy do zbadania relacji między kategoriami. Na podstawie analizy skupień i wizualizacji współwystępowania kategorii za pomocą Gephi przeszło dwustu tysięcy syntagm wyróżniliśmy następujące obszary: sfera publiczna (np. „Aktualności”, „Polityka”), czas wolny (np. „Rozrywka”, „Hobby”), kultura (np. „Muzyka”, „Plotki”), sfera prywatna (np. „Pamiętniki”, „Damskie sprawy”) ¹⁸.

Choć umieszczenie blogów na spektrum między sferą publiczną i prywatną jest spójne z innymi rozpoznaniem na tym polu ¹⁹, uzyskana klasyfikacja nadal wydawała się nam nazbyt tematyczna. Na kolejnym etapie na podstawie analizy treści 322 popularnych polskich blogów zespół zaproponował typologię opartą na wyznacznikach retorycznych, to jest celu komunikacyjnego i motywacji wypowiedzi. Wyróżniliśmy następujące gatunki: diarystyka (opis zdarzeń i subiektywnych doświadczeń), refleksja (subiektywny stosunek do świata wyrażony

¹⁸ Pełny opis wyników w: M. Maryl, K. Niewiadomski, M. Kidawa, *Empirically Generated Typology of Weblog Genres*, „CLCWEB: Comparative Literature and Culture”, [w przygotowaniu].

¹⁹ Zob. A. Morrison, *Blogs and Blogging: Text and Practice*, w: *A Companion to Digital Literary Studies*, red. R. Siemens, S. Schreibman, Oxford: Blackwell 2008, [online:] <http://www.digitalhumanities.org/companionDLS/> [dostęp 27.09.2015]. Zwłaszcza podrozdział *Genres of Blogs*.

w kategoriach uniwersalizujących), krytyka (wypowiedź ekspercka z zakresu danej dziedziny życia publicznego), porada (wskazówki dotyczące działań z życia codziennego), modelowanie (autoprezentacja twórcy jako wzoru do naśladowania), fikcjonalność (opis wydarzeń fikcyjnych)²⁰.

Zaprezentowane na tych etapach typologie to jednocześnie różne sposoby kategoryzacji blogów za pomocą zarówno wyznaczników ilościowych (metadane), jak i jakościowych. Prezentując szczegółową metodologię i kolejne etapy badawcze, skupiam się na technice budowania takich typologii. Uwagi porządkuję względem kolejnych etapów procedury badawczej: dobór materiału, pozyskanie i przetworzenie materiału, analiza i interpretacja.

Dobór materiału

W niniejszym badaniu dobór materiału dotyczył zarówno samych blogów, jak i przeróżnych metadanych. Metadanymi (czyli „danymi o danych”) w kontekście blogów nazywam wszelkie informacje dotyczące tekstów. Jeśli przyjrzeć się raz jeszcze wspomnianemu wcześniej blogowi Segritty, możemy wyróżnić następujące metadane:

²⁰ Ostatni etap badania – wywiady z blogerami – dotyczył strategii autorskich i warsztatu blogera, toteż w trosce o spójność conceptualną tego artykułu nie będzie tu prezentowany. Pełen opis w: A. Buchner, M. Marył, *Nowi literaci. Warsztat twórczy blogerów w kontekście współczesnych przemian kultury literackiej*, w: *Rzeczywistości kulturowe a teksty pisane*, [w przygotowaniu].

tytuł wpisu, nazwy kategorii, do których go przyporządkowano („lifestyle”, „kultura”, „męsko-damskie...” itd.), data publikacji, liczba odsłon, liczba komentarzy, pseudonimy komentujących, liczba udostępnień w serwisach społecznościowych (Facebook, Twitter). Pojęcie metadanych w odniesieniu do piśmiennictwa multimedialnego bliskie jest literaturoznawczemu terminowi „parateksty”, wykorzystywanym na oznaczenie wszystkich elementów pozatekstowych utworu. Do metadanych w piśmiennictwie multimedialnym możemy zaliczyć także wszelkie dane o tekstach, które mogą uzyskać badacze po wstępnym ich przetworzeniu (o czym jeszcze będzie mowa): na przykład liczba słów, długość zdań, iloraz typu i egzemplarzy (*type-token ratio*), liczba rzeczowników konkretnych, słów nacechowanych emocjonalnie i tym podobne.

W wypadku doboru metadanych kierujemy się kryteriami reprezentatywności, jakości i dostępności. A więc, po pierwsze, dane muszą być reprezentatywne dla badanej populacji w znaczeniu zwykle używanym w naukach społecznych. Chodzi zatem o taki dobór próby, by zebrać możliwie najpełniejszy wybór możliwych typów blogów. Po drugie, pod pojęciem jakości rozumiem tu zbieranie danych, które uznamy za przydatne pod kątem badania. Na przykład do badania kategorii wykorzystanych do opisywania blogów w polskiej blogosferze wykorzystaliśmy dane z największych polskich serwisów blogerskich. Dobór próby polegał na identyfikacji polskich platform blogerskich (lista obejmowała 62 platformy w 2013 roku),

odrzućcenie serwisów tematycznych (w trosce o utrzymanie pewnego poziomu ogólności kategoryzacji) i tych, które nie prowadziły interesujących nas klasyfikacji. Po trzecie, dane muszą być dostępne dla badaczy – niektóre z metadanych, jak statystyki odwiedzin, są często ukrywane. Na przykład prezentowane badania syntagmy oparliśmy na materiale z serwisu Bloog.pl, ponieważ tam tylko publikowano listy blogów z przypisanymi do nich wiązkami kategorii. Kryterium dostępności wyznacza czasem kierunek badania, choć czasem prowadzi do pewnych niespodzianek. W czasie badania Bloog.pl ogłaszał, iż w serwisie znajduje się przeszło milion (!) blogów, co z grubsza oznaczałoby, iż 4 procent spośród około 25 milionów użytkowników w Polsce nie tylko bloguje, lecz także robi to w serwisie Bloog.pl. Skuszeni tak wielką próbą i dostępnością metadanych pobraliśmy listy blogów z poszczególnych kategorii (w sumie 55 931 stron zawierających dane 1 116 934 blogów). Niestety niedługo się cieszyliśmy tą próbą – po usunięciu powtórzeń okazało się, że unikalnych blogów jest tylko (sic!) nieco ponad dwieście tysięcy. Co więcej, w miarę postępu badania próba topniała w oczach, o czym za chwilę.

Podobne kryteria stosowaliśmy w przypadku doboru blogów do analiz szczegółowych. Dobór próby w bardzo ciekawy sposób zmusił nas do zoperacjonalizowania pojęcia bloga. Dobierając metadane, za blog uznawaliśmy wszystkie teksty opatrzone tym mianem w serwisach. W przypadku materiału do interpretacji sprawa była

trudniejsza. Powróćmy do historii „milion” blogów. Z listy 200 tysięcy blogów wylosowaliśmy 300 blogów do analiz szczegółowych. Szybko się okazało, że analizy uniemożliwia fakt, iż większość z nich to albo blogi testowe, albo wprowki nastolatków, którzy chcieli w ten sposób utrzymać kontakt z rówieśnikami, ale rezygnowali po dwóch, trzech wpisach. Sporo było także witryn wygenerowanych automatycznie lub mechanicznie na potrzeby pozycjonowania (umieszczanie linków do innych portali). Zjawisko blogów efemerycznych i wątpliwych nazwaliśmy „efektem Turbota” na cześć bloga Turbot (<http://turbotryba.bloog.pl/>), świetnie ilustrującego ten problem: na pustynnym szablonie (kolor żółty i wielbłądy idące przez pustynię) założono blog, składający się z jednego wpisu z 19 października 2011, który zawiera definicję turbota skopiowaną z Wikipedii oraz linki do stron internetowych producenta krerek kuchennych i dwóch firm polonijnych z Wielkiej Brytanii (przesyłki do kraju i remonty). Nieusuwanie takich blogów z serwisu (wszak zwiększają ogólną liczbę blogów, którą serwis może się pochwalić), utrudnia badanie.

Powstało zatem pytanie, czym tak naprawdę jest blog, lub raczej: ile wpisów powinien mieć „prawdziwy” blog? Powróciliśmy do próby i przeanalizowaliśmy ją pod kątem liczby wpisów. Okazało się, że przeszło 1/3 próby (70 799 blogów, 35 procent) zawiera tylko jeden wpis, a blisko połowa (96 304, 48 procent), nie więcej niż dwa. Gdyby arbitralnie założyć, że „prawdziwy” blog ma więcej niż

pięć wpisów, w próbie z Bloog.pl zostałyby 68 887 blogów (34 procent), jeśli poprzeczkę podniesiemy do 10 wpisów, zostanie nam 43 679 blogów (22 procent), jeśli do 20 – 24 413 (12 procent), a już tylko 9305 (5 procent) ma wpisów więcej niż 3989 blogów (2 procent) przekracza 100 notek (to byłby rozmiar odpowiadający średniej wielkości książki), a 88, czyli niecałe pół promila (0,04 procent), przekroczyło 1000 wpisów. To mało, biorąc pod uwagę, że zaczynaliśmy od miliona.

Problem reprezentatywności w dość ciekawy sposób spleta się tu zatem z problemem jakości. Możemy oczywiście wylosować blogi do analizy na przykład spośród 2 procent, które przekroczyły 100 wpisów, ale czy jest to właściwe kryterium doboru blogów? Wszak cykliczne blogi obrazkowe albo filtry mogą się tu wysunąć na prowadzenie. Można sobie też wyobrazić blog w miarę nowy, ale prowadzony regularnie przy udziale publiczności, który ma dopiero kilkadziesiąt wpisów, ale się rozwija. Zespół podjął zatem decyzję o odstąpieniu od losowania próby na rzecz doboru celowego²¹. Uznaliśmy, iż interesują nas te blogi, które faktycznie funkcjonują, autorzy je rozwijają, a czytelnicy odwiedzają i komentują. Na próbę złożyły się zatem trzy grupy blogów: (1) najpopularniejsze blogi z serwisów blox.pl, blog.pl, blog.interia.pl i bloog.pl, (2) blogi wymienione w branżowych rankingach

²¹ Por. Ch.F. Nachmias, D. Nachmias, *Metody badawcze w naukach społecznych*, przeł. E. Hornowska, Poznań: Zysk i S-ka 2001, s. 199.

i raportach o blogosferze²², (3) finaliści i laureaci konkursu Blog Roku (edycja 2012). Łącznie zebraliśmy w ten sposób 322 blogi. Do zalet tej procedury z pewnością można zaliczyć to, iż uzyskaliśmy ciekawą, zróżnicowaną próbę, która pozwoliła na przeprowadzenie analizy treści. Do wad – nadreprezentację w tej próbie blogów kulinarnych i poświęconych mediom społecznościowym. To pierwsze wynika zapewne z faktu, iż blogi kulinarne należą do najczęściej odwiedzanych i pojawiają się wysoko w rankingach, a wykorzystaniem raportów i rankingów poświęconych marketingowi w internecie można tłumaczyć nadreprezentację blogów „branżowych”.

Pobranie materiału

Kolejny etap procesu badawczego, co należy podkreślić z całą mocą, bywa bagatelizowany lub zwyczajnie pomijany przez badaczy. Jakiś czas temu otrzymałem maila od

²² Korzystaliśmy z raportu: T. Baran, A. Miotk, *Blogerzy w Polsce 2013. Znajomość – wizerunek – znaczenie*, Ogólnopolski Panel Badawczy ARIADNA, luty 2013, [online:] [http://pbi.org.pl/aktualnosci/Blogerzy%20w%20Polsce%202013%20\(18kwi13\).pdf](http://pbi.org.pl/aktualnosci/Blogerzy%20w%20Polsce%202013%20(18kwi13).pdf) [dostęp: 27.09.2015] oraz następujących rankingów: tw, *Top polskich blogów: na czele Wojciechowska, Kotlet.tv, Pani Basia, Palikot, Czubaszek i Mucha*, [online:] <http://www.wirtualnemedial.pl/arttykul/top-polskich-blogow-na-czele-wojciechowska-kotlet-tv-pani-basia-palikot-czubaszek-i-mucha> [dostęp 17.02.2016]; <http://www.kominek.in/2014/01/ranking-najbardziej-wplywowych-blogerow-w-2013-roku/>; I. Patro, *Oto których blogerów cenimy i dlaczego*, [online:] <http://ilonapatro.com/2013/12/oto-ktorych-blogerow-cenimy-i-dlaczego/#more> [dostęp 17.02.2016].

zrozpaczonej badaczki, zajmującej się analizą wielkiego portalu fanowskiego, który nieoczekiwanie zniknął z sieci. I jak tu prowadzić badania, jak zweryfikować istniejące rozpoznania, jak żyć? Sformułujmy podstawową zasadę badań internetowych: badany materiał należy pobrać i zapisać w bezpiecznym miejscu. Tylko taka procedura gwarantuje spokojne przeprowadzenie badań i ewentualny powrót do materiału po jakimś czasie. Powodów jest wiele: strona zniknie, treść ulegnie zmianie, struktura zostanie zmodyfikowana. Przykładowo niedługo po rozpoczęciu naszego badania serwis interia.pl zrezygnował ze stosowania kategorii do oznaczania swoich blogów. Szczęśliwie wszystkie kategorie z tego serwisu mieliśmy już zapisane.

Oczywiście w sytuacjach kryzysowych badacze mogą podierać się archiwami internetowymi, jak WaybackMachine witryny archive.org, Google cache czy Coral Content Distribution Network. Są to jednak archiwa wybiórcze i nieregularnie aktualizowane, zwłaszcza w odniesieniu do polskiego internetu. Należy je zatem traktować wyłącznie jako ostatnią deskę ratunku (tamże skierowałem wspomnianą badaczkę).

Pobieranie metadanych może przebiegać jedno- lub dwuetapowo. W tym drugim przypadku najpierw pobieramy cały blog, a następnie dokonujemy ekstrakcji metadanych (o czym będzie mowa w dalszej części artykułu). W tym miejscu skupiam się wyłącznie na bezpośrednim uzyskiwaniu danych. Po pierwsze można pobrać metadane

„ręcznie”, czyli skopiować lub zapisać dostępne zestawienia danych, jak na przykład listę kategorii blogowych czy listę blogów z serwisu bloog.pl²³. Oczywiście w sytuacji, gdy taka lista jest długa (przypomnijmy, iż w przypadku Bloog.pl liczyła sobie 55 931), ręczne pobieranie nie ma sensu i należy wspomóc się oprogramowaniem do pobierania stron, o którym piszę na przykładzie blogów. Drugą metodą jest pobranie materiału przy wykorzystaniu oprogramowania do pozyskiwania danych z sieci (*web scraping*). Lista takich aplikacji jest długa i nie będziemy ich tu szczegółowo omawiać. Pod względem prowadzonych tu badań mogą nas zainteresować dwa typy narzędzi: aplikacje do pozyskiwania konkretnych, ustrukturuowanych danych lub uzyskiwania listy linków dostępnych na stronie. Do tego pierwszego zadania można wykorzystać takie aplikacje, jak Kimono, Data Toolbar, import.io czy Scrapy²⁴. Idea jest za każdym razem podobna – określamy typ danych, który nas interesuje (np. tytuł wpisu, data i liczba komentarzy) i pobieramy dane z określonego zakresu (w zależności od ustawień i opcji aplikacji). Dane udostępniane są w formie tabelarycznej

²³ Przykładowa strona katalogu wygląda tak: Bloog.pl, *Katalog blogów*, [online:] <http://bloog.pl/kid,221,knazwa,Aktualno%B-6ci,katalog.html> [dostęp 17.02.2016].

²⁴ Witryny opisywanych aplikacji: Kimonolabs, [online:] <https://www.kimonolabs.com/> [dostęp 27.09.2015]; Datatoolbar, [online:] <http://datatoolbar.com/> [dostęp 27.09.2015]; Import, [online:] <https://import.io/> [dostęp 27.09.2015]; Scrapy, [online:] <http://scrapy.org/> [dostęp 27.09.2015].

w różnych formatach (np. csv). Na podstawie tych metadanych z blogu można na przykład przeanalizować charakter wspólnoty budowanej wokół bloga, badając, jakim zainteresowaniem cieszą się poszczególne wpisy na różnych blogach²⁵. W drugim przypadku możemy wykorzystać oprogramowanie do analizy struktury strony (np. Xenu Link Sleuth²⁶), które ujawni linki wewnętrzne i prowadzące poza stronę internetową. Pozwala to zbudować sieć powiązań między stronami internetowymi (tym się jednak w tym artykule nie zajmujemy). Problemy pojawiają się często przy witrynach bardziej skomplikowanych strukturalnie. Należy też dodać, iż w większości przypadków narzędzie trzeba ręcznie skalibrować (wskazać odpowiednie typy danych do pobrania). Jest to jednakże dość łatwa i szybka metoda pobrania dużej ilości danych.

W przypadku pobierania blogów, czyli całych stron internetowych, najlepiej sprawdzał się w tym badaniu HTTrack Website Copier²⁷ – program, który sam może przeanalizować strukturę strony internetowej i stworzyć jej kopię lub też zapisać tylko linki wskazane przez badacza. W przypadku stron budowanych dynamicznie (np.

²⁵ Por. M. Maryl, *Blog jako „dziennik elektroniczny” ...*, dz. cyt., s. 101–104. Zwłaszcza tabele 1 i 2.

²⁶ Xenu Link Sleuth, [online:] <http://home.snafu.de/tilman/xenulink.html> [dostęp 27.09.2015].

²⁷ HTTrack Website Copier, [online:] <https://www.httrack.com/> [dostęp 27.09.2015].

w języku PHP) łatwo doprowadzić do zapętlenia programu, który pobiera te same strony w różnych konfiguracjach – można tego uniknąć, stosując różne filtry lub dostarczając listę adresów URL do pobrania (można ją np. uzyskać wstępnie za pomocą wspomnianego już programu Xenu Link Sleuth). Niestety program przeważnie należy dostosować bezpośrednio do konkretnego formatu struktury strony (np. Wordpress) i często odbywa się to na drodze prób i błędów. Na szczęście program jest dość plastyczny i za pomocą różnych ustawień i filtrów da się go dostosować do większości stron (pomocne jest rozbudowane forum projektu)²⁸. Z pobieraniem blogów wiąże się też problem miejsca na dysku i użycia danych. Często po pobraniu blogu należy dokonać analizy pobranej zawartości i oddzielić same wpisy innych materiałów lub list wpisów. Jeżeli interesuje nas wyłącznie tekst bloga, na przykład do przeprowadzenia analiz stylometrycznych, to często możemy go pobrać za pomocą opisywanego wcześniej oprogramowania do metadanych (np. Kimono)²⁹.

²⁸ Tamże – forum, [online:] <http://forum.httrack.com/> [dostęp 27.09.2015].

²⁹ W ramach naszego projektu powstał we współpracy z Politechniką Wrocławską i konsorcjum CLARIN-PL prototyp narzędzia do pobierania całych blogów wraz z metadanymi, który zaprezentowano w 2014 na konferencji *Practical Applications of Linguistic Corpora* w Łodzi. Prace trwają i narzędzie nie jest jeszcze niestety dostępne.

Przetworzenie materiału

Kolejnym etapem jest przetworzenie materiału do formy potrzebnej w danym badaniu lub wymaganej przez program analityczny, którym chcielibyśmy się posłużyć. Generalnie można założyć, że w przypadku metadanych powinna być to tabela, którą można otworzyć w arkuszu kalkulacyjnym (i ew. przekonwertować na inne formaty lub zaimportować do programów analitycznych). Dane w arkuszu kalkulacyjnym można dalej przetwarzać za pomocą formuł (np. skracając adresy URL różnych podstron) lub prostych makr czy skryptów VBA. Aby odwołać się do przykładu, w naszym badaniu dokonywaliśmy w ten sposób obliczeń frekwencyjnych występowania syntagm – z listy syntagm wyciągaliśmy poszczególne typy za pomocą komendy „Filtruj dane” w MS Excel, a następnie zliczaliśmy ich występowanie przy wykorzystaniu formuły „Licz jeżeli”.

Przy interpretacji całych blogów najlepiej sprawdza się format html lub zapis pdf (by zachować strukturę layoutu i treść ładowaną dynamicznie z innych plików, np. obrazki). Do analiz tekstowych z pewnością przyda się bardziej prosty plik tekstowy (świetnie się do tego nadaje program HTMLAsText³⁰). Tak uzyskane pliki tekstowe można dalej przetwarzać, korzystając ze skryptów

³⁰ NirSoft, [online:] <http://www.nirsoft.net/utils/htmlastext.html> [dostęp 27.09.2015].

i makr, zliczając na przykład liczbę słów, występowanie określonych wyrazów i tym podobne.

Analiza materiału

W tej części tekstu skupię się na trzech typach analiz: interpretacji metadanych, relacji między nimi oraz na komputerowo wspomaganej interpretacji tekstu. Omawiam procedurę badawczą w miarę szczegółowo, by można było ją wykorzystać do badania innego materiału. We wnioskach powrócę do omówienia relacji między tymi metodami przy tworzeniu typologii piśmiennictwa multimedialnego.

Interpretacja metadanych w naszym badaniu dotyczyła analizy klasyfikacji zastanej, czyli kategorii, którymi posługują się serwisy blogowe. Po stworzeniu listy 287 kategorii (w tym 162 podkategorii) przystąpiliśmy do kodowania w duchu teorii ugruntowanej – przyglądając się poszczególnym kategoriom (i samym blogom klasyfikowanym w tych kategoriach) próbowaliśmy zrozumieć ich znaczenie i pogrupować je na zasadzie podobieństwa³¹. Uzyskane w ten sposób grupy kategorii były analizowane, dyskutowane, dzielone i redukowane na zebraniach zespołu, aż ostatecznie przyjęto pięć kategorii wraz z definicjami (omawianymi wcześniej).

³¹ Zob. M. Lonkila, *Grounded theory as an emerging paradigm for computer-assisted qualitative data analysis*, w: *Computer-Aided Qualitative Data Analysis: Theory, Methods and Practice*, red. U. Kelle, London: Sage 1995, s. 41–51.

Następnie opracowaliśmy prosty arkusz kalkulacyjny, w którym trzech koderów przypisywało do każdej kategorii jeden z pięciu kodów (dla celów badawczych stosowaliśmy kod główny i drugi wybór kodu dodatkowego, który pomijam w tym opisie). Zgodność kodowania obliczyliśmy dzięki ReCal3, uzyskując zgodność rzędu 86 procent (κ 0,722), co należy uznać za zadowalający wynik³². Zastosowana procedura pozwoliła ukazać, iż większość (70 procent) kategorii stosowanych do klasyfikacji blogów w dużych serwisach ma charakter tematyczny i trudno zatem wykorzystywać zastane kategorie jako podstawę do budowy klasyfikacji gatunkowej.

Relacje między metadanymi analizowaliśmy, żeby zobaczyć, co syntagmy (wiązki kategorii wykorzystywanych do opisu blogów) mogą nam powiedzieć o samych blogach. Postanowiliśmy posłużyć się w tym celu wizualizacją eksploracyjną, czyli próbą przestrzennego oddania relacji współwystępowania między elementami zbioru³³.

³² Program ReCal3 można pobrać tutaj: ReCal3 – Download, [online:] <http://dfreelon.org/utils/recalfront/recal3/> [dostęp 27.09.2015]. Artykuł tłumaczący działanie narzędzia: D. Freelon, *ReCal: Intercoder reliability calculation as a web service*, „International Journal of Internet Science” 2010, nr 5 (1), s. 20–33, [online:] http://www.ijis.net/ijis5_1/ijis5_1_freelon_pre.html [dostęp 27.09.2015].

³³ Problematykę wizualizacji eksploracyjnej jako metody analitycznej w naukach humanistycznych i społecznych dobrze przybliżają następujące publikacje: *Mastering the Information Age. Solving Problems with Visual Analytics*, red. D. Keim, J. Kohlhammer, G. Ellis, F. Mansmann, Bad Langensalza: Druckhaus Thomas

Wizualizacja zdaje się dość dobrze odpowiadać postulatom CEDA, łącząc ilościowe podstawy (twarde dane liczbowe) z procesem interpretacji wizualnej postaci relacji między nimi. Warto podkreślić, iż zwłaszcza w przypadku wizualizacji za pomocą Gephi ważny jest nie tyle punkt dojścia (np. taki wykres, jaki prezentuję na rycinie 9.1), ile sam proces przyglądania się danym, dynamicznego zestawiania elementów i interpretowania rezultatów na bieżąco.

Do pierwszej wizualizacji w formie dendrogramu zastosowaliśmy hierarchiczną analizę skupień. W arkuszu kalkulacyjnym przekształciliśmy każdą syntagmę w osobny przypadek, który rozpisaliśmy za pomocą kategorii nominalnych. Czyli syntagma „Polityka-Praca-Psychologia” została opisana jako przypadek, w którym przyznajemy 1 w tych trzech kategoriach i zera przy pozostałych 42. Tak przygotowany arkusz kalkulacyjny (w formie pliku csv) zaimportowaliśmy do programu SPSS. Ze względu na fakt, iż wykorzystaliśmy dane nominalne, zastosowaliśmy

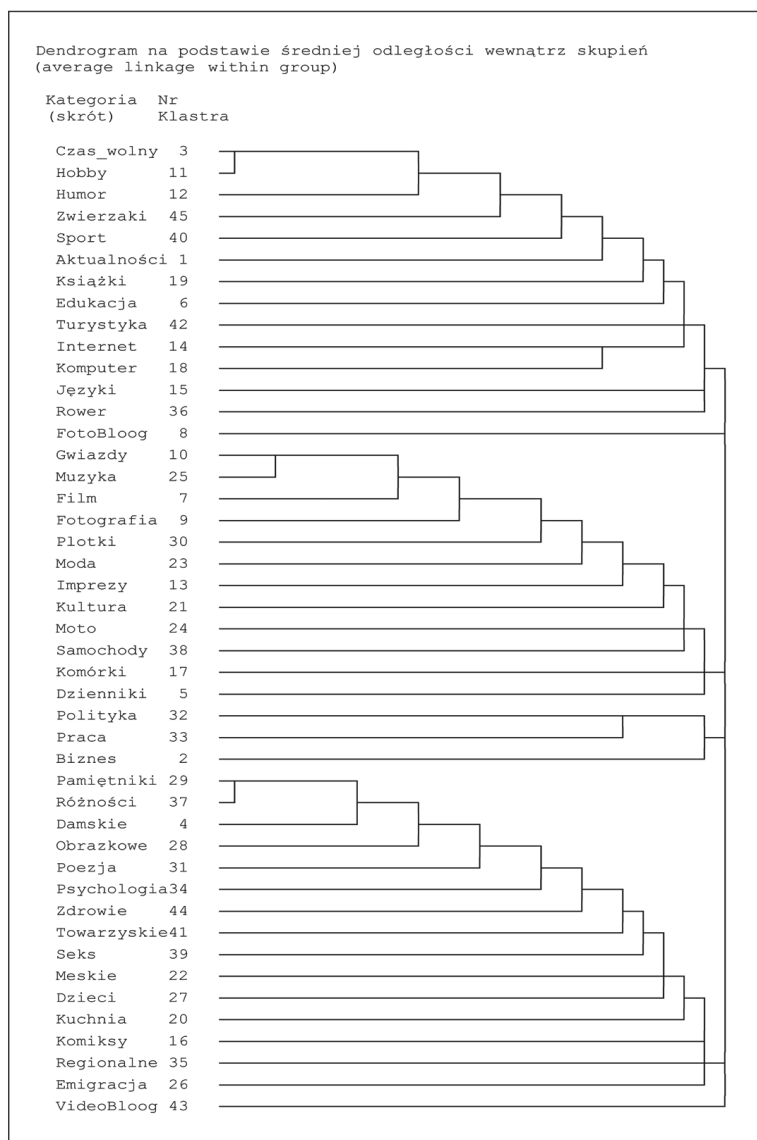
Müntzer GmbH 2010, [online]: <http://www.vismaster.eu/wp-content/uploads/2010/11/VisMaster-book-lowres.pdf> [dostęp 27.09.2015]; M. B. Paradowski, *Wizualizacja danych – dużo więcej, niż prezentacja*, [online:] <http://publikacje.ils.uw.edu.pl/publication/view/polski-wizualizacja-danych-duzo-wiecej-niz-prezentacja/> [dostęp 27.09.2015]; tegoż, *Dekalog analityka danych i infografika – quid, cur, quomodo*, w: *Wizualizacja wiedzy. Od Biblia Pauperum do hipertekstu*, red. M. Kluza, Lublin: Portal Wiedza i Edukacja 2011, [online:] <http://wiedzaiedukacja.eu/wp-content/uploads/2012/03/wizualizacijawiedzy.pdf> [dostęp 27.09.2015].

algorytm analizy skupień Russela i Rao oraz opcję powiązań wewnątrzgrupowych w programie SPSS jako metody porządkującej grupy (clustery), żeby uzyskać możliwie najmniejsze dystanse między elementami grupy³⁴. Wyniki prezentujemy na rycinie 9.1.

Widać tu wstępne uporządkowanie w grupy, które można zinterpretować w następujący sposób:

1. Sfera publiczna i rozrywka („Czas wolny”, „Hobby”, „Humor i Rozrywka”, „Zwierzaki”, „Sport”, „Aktualności”, „Książki”, „Edukacja”, „Turystyka”, „Internet”, „Komputer”, „Języki obce”, „Rower”).
2. Rozrywka i kultura („Gwiazdy”, „Muzyka”, „Film”, „Fotografia”, „Plotki”, „Moda”, „Imprezy”, „Kultura”, „Motoryzacja”, „Samochody”, „Komórki”, „Dziennik budowy”).
3. Sfera Publiczna („Polityka”, „Praca”, „Biznes”).
4. Sfera prywatna („Pamiętniki”, „Różności”, „Damskie sprawy”, „Obrazkowe”, „Poezja”, „Psychologia”, „Zdrowie i uroda”, „Towarzyskie”, „Seks i erotyka”, „Męskie sprawy”, „Nasze dzieci”, „Kuchnia”, „Komiksy”, „Regionalne”, „Na emigracji”, „VideoBloogi”).

³⁴ Zob. H. Finch, *Comparison of distance measures in cluster analysis with dichotomous data*, „Journal of Data Science” 2005, nr 3.1, s. 87; P.-N. Tan, M. Steinbach, V. Kumar, *Introduction to Data Mining*. Boston: Pearson Addison Wesley 2005, s. 519–524.



RYCINA 9.1.

Dendrogram przedstawiający pogrupowane kategorie

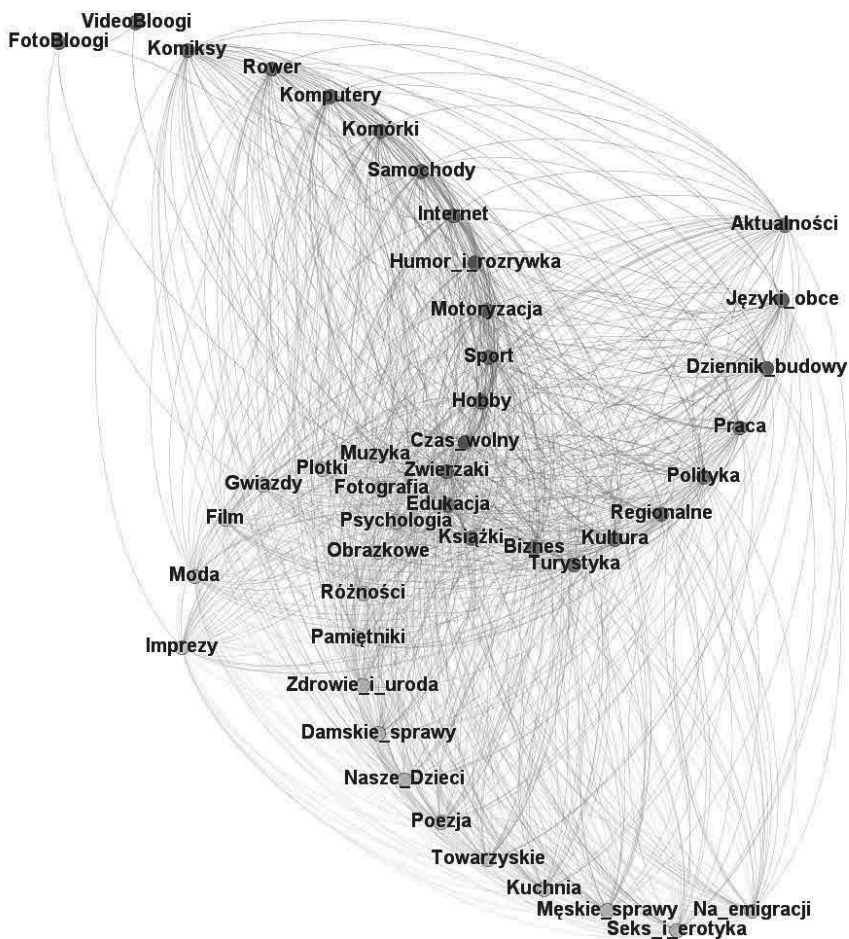
Źródło: opracowanie własne.

Jeżeli przyjrzeć się lepiej rycinie 9.1, widać, iż te podziały wcale nie zawsze są takie oczywiste. Na przykład dlaczego „Motoryzacja” należy do grupy 2, a „Komiksy” do 4? Żeby lepiej zrozumieć relacje między tymi elementami, skorzystaliśmy z innej metody wizualizacji, oferowanej przez program Gephi³⁵.

Gephi służy do wizualizacji relacji (krawędzi) między dwoma elementami (węzłami). Wizualizacja wymagała odpowiedniego przekształcenia syntagma w arkuszu kalkulacyjnym w listę par (np. syntagma „Polityka-Praca-Psychologia” została przekształcona w trzy pary „Polityka-Praca” i „Praca-Psychologia” i „Polityka-Psychologia”), a następnie zaimportowana w formie csv do Gephi jako sieci nieukierunkowanej (*undirected*), to znaczy takiej, w której kierunek relacji nie jest istotny. Gephi zlicza wagę każdej krawędzi, czyli siłę relacji między poszczególnymi węzłami (im częściej powtarzały się dane połączenia, tym większą wagę będą miały). W rezultacie uzyskaliśmy sieć 45 węzłów (czyli kategorii) i 1799 krawędzi (czyli powiązań między nimi). Do analizy wykorzystaliśmy dostępny w Gephi algorytm Louvain pozwalający na identyfikację grup³⁶. Algorytm pozwolił wskazać 4 grupy ($Q = 0,124$), które widzimy na rycinie 9.2.

³⁵ Gephi, [online:] <http://gephi.github.io/> [dostęp 27.09.2015]. Na stronie można znaleźć bardzo dużo materiałów informacyjnych i praktyczne samouczki.

³⁶ Zob. V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, *Fast unfolding of communities in large networks*, „Journal of Statistical



RYCINA 9.2.

Relacje między kategoriami. Wizualizacja za pomocą Gephi

Źródło: opracowanie własne.

Do wizualizacji wykorzystaliśmy layout osi promienistych (*Radial Axis layout*). Dzięki tym zabiegom łatwiej wyróżnić grupy odpowiadające sferze publicznej (z prawej: „Aktualności”, „Polityka”, „Praca” itd.), prywatnej (na dole: „Psychologia”, „Pamiętniki”, „Towarzyskie” itd.), rozrywce (na górze: „Czas wolny”, „Zwierzaki”, „Humor i rozrywka” itd.) i kulturze (z lewej: „Muzyka”, „Plotki”, „Gwiazdy” itd.).

Choć prezentowana tu typologia zdaje się lepiej porządkować materiał niż ta, którą uzyskaliśmy w pierwszym badaniu, to nadal nie udało się nam uciec od kryterium tematycznego. Dodatkowo nasze wątpliwości budził sam fakt wykorzystania takiego materiału jak kategorie dobierane przez autorów przy zakładaniu blogów: nie mamy pewności, na ile był to wybór świadomy, a na ile mechaniczny i niewiele znaczący. Co więcej, tylko 35 kategorii istniało w serwisie od początku (przed 2005/2006). W 2006 dodano jedną („Gwiazdy”), w 2007 trzy kolejne („Plotki”, „Polityka”, „Zwierzaki”), dwie w 2007 („Dziennik budowy”, „Biznes”), jedną w 2009 („Moda”) i po dwie w 2010 („Na emigracji”, „Nasze dzieci”) i 2011 („FotoBloogi”, „VideoBloogi”)³⁷. Ta ewolucja może tłumaczyć mniejszą popularność tych ostatnich kategorii. Wreszcie kategorie – choć ewoluują

³⁷ Rekonstrukcji dokonano przy pomocy Wayback Machine na stronie Internet Archive, [online:] <http://www.archive.org/> [dostęp: 29.09.2015].

– stanowią formę konceptualną narzuconą przez operatorów serwisu. Poszukując innego sposobu konceptualizacji materiału, zaplanowaliśmy badanie interpretacyjne, w którym uporządkowaliśmy materiał według innych wyznaczników.

Interpretacja tekstu wspomagana komputerowo została przeprowadzona na materiale 322 popularnych blogów (o których doborze pisałem). Proces badawczy przebiegał tu w podobny sposób jak w przypadku analizy kategorii w pierwszym badaniu. Różnica polegała na zmianie podejścia interpretacyjnego i wykorzystaniu programu Atlas.ti do komputerowego wspomagania interpretacji w duchu CAQDAS (*Computer Assisted Qualitative Data Analysis*).

Przyjęliśmy definicję Carolyn Miller, traktującą gatunek mowy w sposób pragmatyczny i funkcjonalny, jako czynność społeczną (*social action*). Według badaczki każdy gatunek staje się reprezentacją pewnej praktyki, oddającą daną sytuację oraz motywacje aktorów społecznych. Gatunek, powiada Miller, „staje się w tej perspektywie czymś więcej, niż tylko bytem formalnym; staje się pragmatycznym, w pełni retorycznym punktem wspólnym intencji i skutku, częścią czynności społecznej”³⁸. Założyliśmy zatem, iż interpretacja tekstu powinna zakładać następujące

³⁸ C. R. Miller, *Genre as Social Action*, w: *Genre and the New Rhetoric*, red. A. Freedman, P. Medway, London: Taylor & Francis 1994, s. 21.

aspekty: cel komunikacyjny (intencja), przedmiot wypowiedzi (opisany na spektrum między sferą publiczną a prywatną), rolę nadawcy, dyskursywną dominantę blogu oraz strategię narracyjną mającą zapewnić blogom spójność. Wszystkie te wyznaczniki wzięto pod uwagę przy pracy analitycznej.

Część zespołu korzystała z programu Atlas.ti, który pozwala na opatrywanie treści pobranych blogów odpowiednimi kodami, a następnie przeglądanie notatek według konkretnych kodów. Procedura wygląda tu następująco – odpowiednimi kodami opatruje się miejsca w tekście, a następnie analizuje fragmenty przyporządkowane do poszczególnych kodów w trybie omawianym już przy badaniu numer 1.

Wyróżnione gatunki (omówione na wstępie do tego tekstu) zostały uzgodnione podczas spotkań zespołu. To ważny element procesu, podczas którego wielokrotnie wracamy do tekstu i dookreślamy definicje. Członkowie zespołu czytali przypadającą im partię blogów, przy czym wymieniali się uwagami i prezentowali najbardziej wyraziste przykłady. Następnie zespół analizował dalsze partie materiału i dokonywał weryfikacji definicji i kolejnych uzgodnień. Warto podkreślić – co jest głównym wyzwaniem metodologicznym – iż potraktowano je jako swoiste typy idealne czy też prototypy w rozumieniu kognitywistycznym, a zatem pewne modelowe przypadki, do których konkretne blogi się zbliżają, ale mogą przy tym realizować cechy innego gatunku.

Konkluzje

Zaprezentowałem tu metodę budowania typologii gatunków piśmiennictwa multimedialnego, bazując na materiale empirycznym. Wszystkie trzy typologie są polityczne, to znaczy oparte nie tyle na przynależności elementu do danej grupy na podstawie wszystkich jego własności, ile na bliskości i podobieństwie między elementami³⁹. W badaniu pierwszym i trzecim zbudowaliśmy kategorie w ramach indukcji inspirowanej teorią ugruntowaną, w drugim zaś badaniu posłużyliśmy się ilościowym zestawieniem relacji między elementami.

Poświęćmy chwilę na analizę relacji między tymi technikami i konsekwencjami typologii. Typologie w dwóch pierwszych badaniach oparliśmy na istniejących konceptualizacjach (typologie naturalne), czyli *de facto* pracowaliśmy na typologiach tworzonych przez samych użytkowników i serwisy blogowe. W pierwszym badaniu stworzyliśmy konceptualną typologię kategorii z portali, by pokazać, jakiego typu wyznaczniki są najistotniejsze w blogosferze. Okazało się, iż dominują kryteria tematyczne, inne cechy zaś (w tym formalne i gatunkowe) schodzą na plan dalszy. W drugim – pokazaliśmy relacje między tymi kategoriami na podstawie faktycznych użyc tych kategorii przez samych blogerów. Udało się

³⁹ Zob. K. D. Bailey, *Typologies...*, dz. cyt., s. 7.

uzyskać cztery grupy blogów, odpowiadające różnym obszarom zainteresowania, co pokrywa się naszym zdaniem z typami zróżnicowania współczesnej oferty prasowej. Trzecie badanie przeprowadziliśmy po to, by uciec od potocznych konceptualizacji, mających za podstawę kryteria tematyczne. Interpretacja blogów pozwoliła wskazać kilka prototypowych gatunków, wyróżnionych ze względu na cel i funkcję wypowiedzi. Udało się nam w ten sposób pokazać ciągłość gatunków blogowych w relacji do gatunków *offline* oraz ukazać części wspólne bez względu na tematykę (np. do gatunku „krytyka” zaliczyliśmy zarówno felietony sportowe, jak i publicystykę ekonomiczną).

Proponowana tu metoda ma charakter autorski i sama stanowi jeden z efektów projektu badawczego. Od początku chcieliśmy zastosować różne narzędzia, by zobaczyć, jakie wyniki można dzięki nim uzyskać oraz ocenić ich przydatność. Formuła tekstu metodologicznego pozwala opowiedzieć ze szczegółami o tym procesie, a także o jego ślepych uliczkach, jak choćby wspomniana już pułapka próby losowej, w którą wpadliśmy. Na koniec warto jeszcze napisać o potencjalnym wykorzystaniu tej metodologii do innych badań. Większość opisywanych tu programów i aplikacji jest darmowa lub posiada darmowe odpowiedniki. Często bywa tak, iż tematyka badania czy własności dostępnego materiału zmuszają do szukania konkretnej aplikacji, która posiadałaby jakąś jedną, wyjątkową funkcję, niezbędną do przeprowadzenia badania.

Warto polecić Digital Research Tools (DiRT) Directory⁴⁰, w którym badacze znajdą ogromny katalog dostępnych narzędzi wraz z opisem zastosowania. Badania internetu mają bowiem wiele z bricolage'u – przykładamy różne narzędzia, by pozwolić naszym danym mówić.

BIBLIOGRAFIA

- Bailey K. D., *Typologies and taxonomies: an introduction to classification techniques*, Thousand Oaks: Sage 1994.
- Balcerzan E., *W stronę genologii multimedialnej*, „Teksty Drugie” 1999, nr 6.
- Baran T., Miotk A., *Blogerzy w Polsce 2013. Znajomość – wizerunek – znaczenie*, Ogólnopolski Panel Badawczy ARIADNA, luty 2013, [online:] [http://pbi.org.pl/aktualnosci/Blogerzy%20w%20Polsce%202013%20\(18kwi13\).pdf](http://pbi.org.pl/aktualnosci/Blogerzy%20w%20Polsce%202013%20(18kwi13).pdf) [dostęp: 27.09.2015].
- Blondel V. D., Guillaume J.-L., Lambiotte R., Lefebvre E., *Fast unfolding of communities in large networks*, „Journal of Statistical Mechanics: Theory and Experiment” 2008, nr 10, [online:] <http://arxiv.org/pdf/0803.0476v2.pdf> [dostęp 27.09.2015].
- Bloog.pl, *Katalog blogów*, [online:] <http://bloog.pl/kid,221,knazwa,Aktualno%B6ci,katalog.html> [dostęp 17.02.2016].
- Buchner A., Maryl M., *Nowi literaci. Warsztat twórczy blogerów w kontekście współczesnych przemian kultury literackiej*, w: *Rzeczywistości kulturowe a teksty pisane*, [w przygotowaniu].

⁴⁰ DiRT – Digital Research Tools, [online:] <http://dirtdirectory.org/> [dostęp 27.09.2015].

- Datatoolbar, [online:] <http://datatoolbar.com/> [dostęp 27.09.2015].
- DiRT – Digital Research Tools, [online:] <http://dirttdirectory.org/> [dostęp 27.09.2015].
- Finch H., *Comparison of distance measures in cluster analysis with dichotomous data*, „Journal of Data Science” 2005, nr 3.1.
- Freelon D., *ReCal: Intercoder reliability calculation as a web service*, „International Journal of Internet Science” 2010, nr 5 (1), s. 20–33, [online:] http://www.ijis.net/ijis5_1/ijis5_1_freelon_pre.html [dostęp 27.09.2015].
- Gephi, [online:] <http://gephi.github.io/> [dostęp 27.09.2015].
- Gumkowska A., *Blogi wobec tradycji diarystycznej. Nowe gatunki w nowych mediach*, w: *Tekst (w) sieci, t. 1: Tekst, język, gatunki*, red. D. Ulicka, Warszawa: Wydawnictwa Akademickie i Profesjonalne 2009.
- Hanauer D., *What we know about reading poetry: Theoretical positions and empirical research*, w: *The Psychology and Sociology of Literature: In Honor of Elrud Ibsch*, red. D. Schram, G. Steen, Amsterdam – Philadelphia: John Benjamins 2001.
- HTTRack Website Copier – forum, [online:] <http://forum.httrack.com/> [dostęp 27.09.2015].
- HTTRack Website Copier, [online:] <https://www.httrack.com/> [dostęp 27.09.2015].
- Import, [online:] <https://import.io/> [dostęp 27.09.2015].
- Internet Archive, [online:] <http://www.archive.org/> [dostęp: 29.09.2015].
- KimonoLabs, [online:] <https://www.kimonolabs.com/> [dostęp 27.09.2015].
- Lonkila M., *Grounded theory as an emerging paradigm for computer-assisted qualitative data analysis*, w: *Computer-Aided Qualitative*

- Data Analysis: Theory, Methods and Practice*, red. U. Kelle, London: Sage 1995.
- Maryl M., *Blog jako „dziennik elektroniczny”*. Analiza genologiczna blogów pisarzy, „Zagadnienia Rodzajów Literackich” 2013, nr 56 (112), z. 2.
- Maryl M., *Konwergencja i komunikacja. Gatunki wypowiedzi na stronach internetowych pisarzy*, „Zagadnienia Rodzajów Literackich” 2012, nr 55 (110), z. 2.
- Maryl M., Niewiadomski K., *Blog = Książka? Empiryczne badanie potocznej kategoryzacji blogów*, „Przegląd Humanistyczny” 2013, nr 4.
- Maryl M., Niewiadomski K., Kidawa M., *Empirically Generated Typology of Weblog Genres*, „CLCWEB: Comparative Literature and Culture”, [w przygotowaniu].
- Maryl M., *Reprint i hipermedialność – dwa kierunki rozwoju literatury ucyfrowionej*, w: *Tekst w sieci, t. 2: Literatura, Społeczeństwo, Komunikacja*, red. A. Gumkowska, Warszawa: Wydawnictwa Akademickie i Profesjonalne 2009.
- Maryl M., *Virtual Communities – Real Readers: New Data in Empirical Studies of Literature*, w: *New Beginnings in the Study of Literature*, red. W. van Peer, J. Auracher, Cambridge: Cambridge Scholars Publications 2008.
- Mastering the Information Age. Solving Problems with Visual Analytics*, red. D. Keim, J. Kohlhammer, G. Ellis, F. Mansmann, Bad Langensalza: Druckhaus Thomas Müntzer GmbH 2010, [online]: <http://www.vismaster.eu/wp-content/uploads/2010/11/VisMaster-book-lowres.pdf> [dostęp 27.09.2015].
- Miller C. R., *Genre as Social Action*, w: *Genre and the New Rhetoric*, red. A. Freedman, P. Medway, London: Taylor & Francis 1994.

- Morrison A., *Blogs and Blogging: Text and Practice*, w: *A Companion to Digital Literary Studies*, red. R. Siemens, S. Schreibman, Oxford: Blackwell 2008, [online:] <http://www.digitalhumanities.org/companionDLS/> [dostęp 27.09.2015].
- Nachmias Ch. F., Nachmias D., *Metody badawcze w naukach społecznych*, przeł. E. Hornowska, Poznań: Zysk i S-ka 2001.
- NirSoft, [online:] <http://www.nirsoft.net/utils/htmlastext.html> [dostęp 27.09.2015].
- Patro I., *Oto których blogerów cenimy i dlaczego*, [online:] <http://ilonapatro.com/2013/12/oto-ktorych-bloggerow-cenimy-i-dlaczego/#more> [dostęp 17.02.2016].
- Paradowski M. B., *Dekalog analityka danych i infografika – quid, cur, quomodo*, w: *Wizualizacja wiedzy. Od Biblia Pauperum do hipertekstu*, red. M. Kluza, Lublin: Portal Wiedza i Edukacja 2011, [online:] <http://wiedzaiedukacja.eu/wp-content/uploads/2012/03/wizualizacjawiedzy.pdf> [dostęp 27.09.2015].
- Paradowski M. B., *Wizualizacja danych – dużo więcej, niż prezentacja*, [online:] <http://publikacje.ils.uw.edu.pl/publication/view/polski-wizualizacja-danych-duzo-wiecej-niz-prezentacja/> [dostęp 27.09.2015].
- ReCal3 – Download, [online:] <http://dfreelon.org/utils/recalfront/recal3/> [dostęp 27.09.2015].
- Rogers R., *Digital methods*, Boston: The MIT Press 2013.
- Scrapy, [online:] <http://scrapy.org/> [dostęp 27.09.2015].
- Segritta, [online:] <https://web.archive.org/web/20150314191513/http://segritta.pl/> [dostęp 10.10.2015].
- Stockwell P., *Poetyka kognitywna. Wprowadzenie*, przeł. A. Skucińska, Kraków: Universitas 2006.
- Sudweeks F., Simoff S. J., *Complementary Explorative Data Analysis. The Reconciliation of Quantitive and Qualitative Principles*, w: *Doing Internet research: Critical issues and methods for examining the Net*, red. S. Jones, Thousand Oaks: Sage 1998.

- Szczepan-Wojnarska A., *Blogi jako forma literacka*, „Pamiętnik literacki” 2006, z. 4.
- Tan P.-N., Steinbach M., Kumar V., *Introduction to Data Mining*, Boston: Pearson Addison Wesley 2005.
- Theisen C. A., Oberlander J., Kirby S., *Systematicity and arbitrariness in novel communication systems*, „Interaction Studies” 2010, nr 11 (1).
- tw, *Top polskich blogów: na czele Wojciechowska, Kotlet.tv, Pani Basia, Palikot, Czubaszek i Mucha*, [online:] <http://www.wirtualnemedial.pl/artukul/top-polskich-blogow-na-czele-wojciechowska-kotlet-tv-pani-basia-palikot-czubaszek-i-mucha> [dostęp 17.02.2016].
- Witosz B., *Genologia lingwistyczna. Zarys problematyki*, Katowice: Wydawnictwo Uniwersytetu Śląskiego 2005.
- Xenu Link Sleuth, [online:] <http://home.snafu.de/tilman/xenulink.html> [dostęp 27.09.2015].